UNIT –V

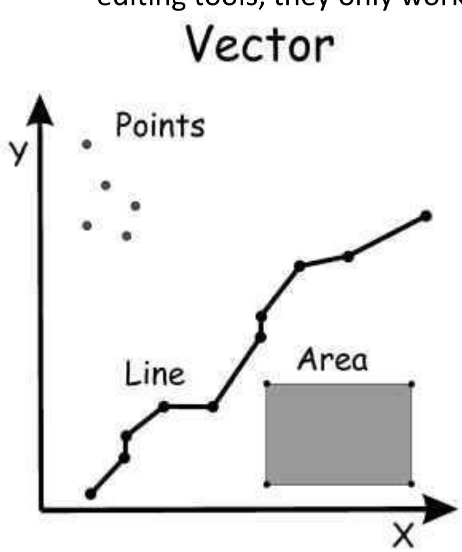
Vector data analysis: Introduction, buffering, map overlay, Distance measurement and map manipulation.

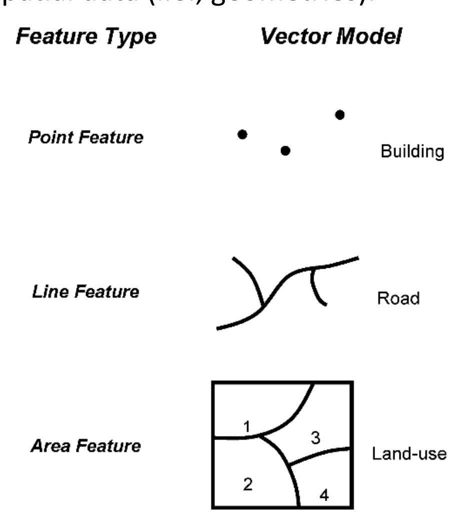
Raster data analysis: Data analysis environment, local operations, neighbourhood operations, zonal operations, Distance measure operations.

Q.1 Explain Vector Data Analysis

VECTOR DATA ANALYSIS

- The scope of GIS analysis varies among disciplines that use GIS.
- GIS users in hydrology will likely emphasize the importance of terrain analysis and hydrologic Modeling, whereas GIS users in wildlife management will be more interested in analytical functions dealing with wildlife point locations and their relationship to the environment.
- The vector data model uses points and their x-. Y-coordinates to construct spatial features of points, lines, and polygons.
- Therefore vector data analysis uses the geometric objects of point, line, and polygon. And the accuracy of analysis results depends on the accuracy of these objects in terms of location and shape.
- Because vector data may be topology-based or non-topological, topology can also be a factor for some vector data analyses such as buffering and overlay.
- A number of analytical tools such as Union and Intersect also appear as editing tools. Although the terms are the same, they perform different functions.
- As overlay tools, Union and Intersect work with both spatial and attribute data. But as editing tools, they only work with spatial data (i.e., geometries).





Q.2 Explain Buffering

Buffering:-

o Based on the concept of proximity buffering creates two areas:
o One area is within a specified distance of select feature
o Other that is beyond.
o The area that is within the specified distance is called the buffer zone.
o Features for buffer may be points, lines or areas.
o Buffering around points creates circular buffer zones.
o Buffering around lines creates a series of elongated buffer zones.
o Buffering around polygons creates buffer zones that extend outward from the polygon boundaries.

Variations in buffering

o The buffer distance or buffer size does not have to be constant; it can vary according to the values of a given field.
o A feature may have more than one buffer zone.
o As an example, a nuclear power plant may be buffered with distances of 5, 10, 15, and 20 miles, thus forming multiple rings around the plant (Figure 1).

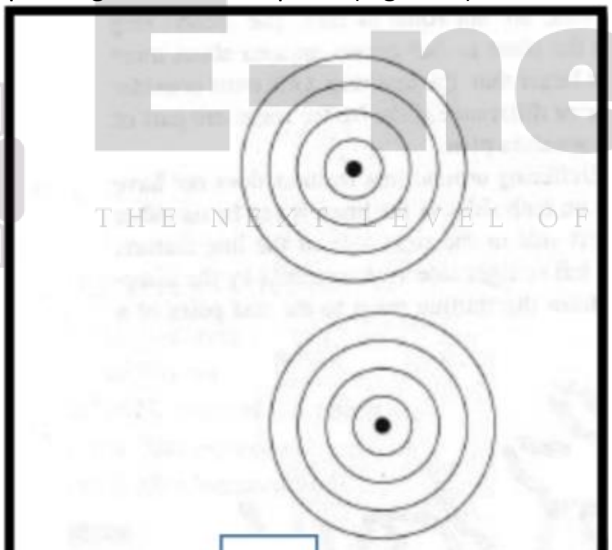


Figure 12 1
Buffering with four rings.

o These buffer zones, although spaced equally from the plant, are not equal in area. The second ring from the plant in fact covers an area about three times larger than the first ring.
o One must consider this area difference if the buffer zones are part of an evacuation plan.
o Buffering around line features does not have to be on both sides of the lines, it can be on either the left side or the right side of the line feature. (The left or right side is determined by the direction from the starting point to the end point of a line).
o Likewise, buffer zones around polygons can be extended either outward or inward from the polygon boundaries.

o Boundaries of buffer zones may remain intact so that each buffer zone is a separate polygon. Or these boundaries may be dissolved so that there are no overlapped areas between buffer zones (Figure 2).
o Even the ends of buffer zones can be either rounded or flat. Regardless of its variations, buffering uses distance measurements from select features to create the buffer zones.

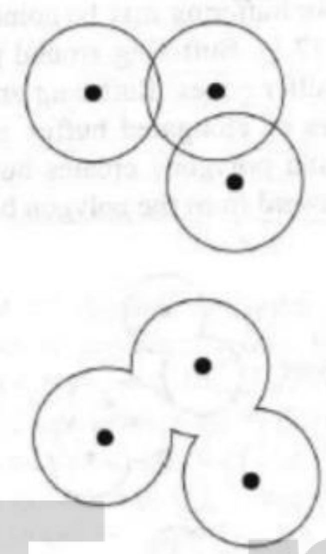


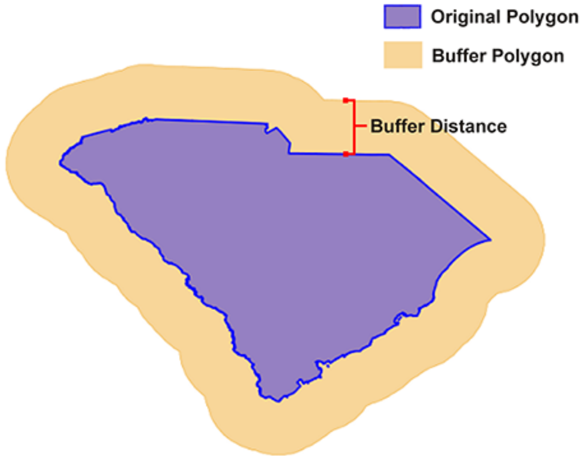
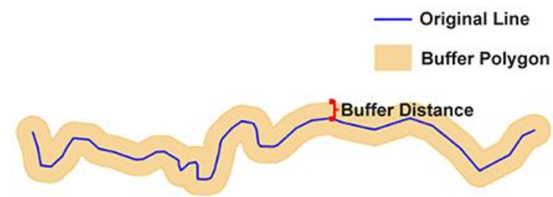
Figure 12.4
Buffer zones not dissolved (top) or dissolved (bottom).

Q.3 Applications of Buffering

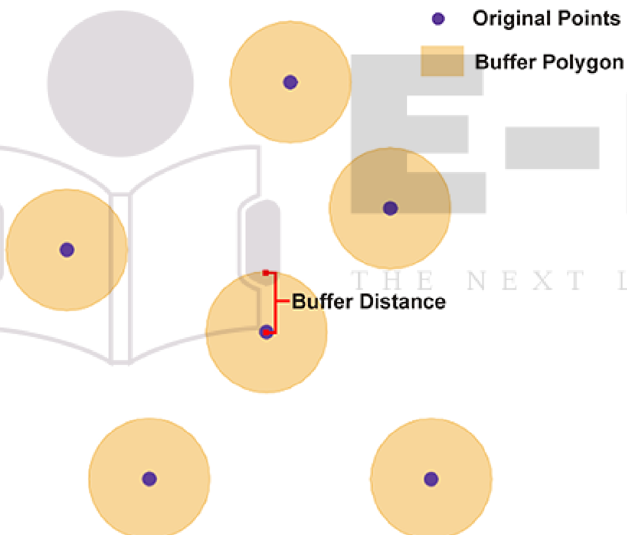
Buffering creates a buffer zone data set, which sets the buffering operation apart from the use of proximity measures for spatial data query Spatial data query using the proximity relationship can select spatial features that are located within a certain distance of other features but cannot create a buffer zone data set.

A buffer zone is often treated as a protection zone and is used for planning or regulatory purposes:

• A city ordinance may stipulate that no liquor stores or pornographic shops shall be within 1000 feet of a school or a church.
• Government regulations may set 2-mile buffer zones along streams to minimize sedimentation from logging operations.
• A national forest may restrict oil and gas well drilling within 500 feet of roads or highways.

Buffer applied to a polygon:	Buffer applied to a line:
	

Buffer applied to a set of points:



Q.4 Explain what is Overlay and its types

Overlay:-

- An overlay operation combines the geometries and attributes of two feature layers to create the output.
- The geometry of the output represents the geometric intersection of features from the input layers.
- Figure 3 illustrates an overlay operation with two polygon layers.

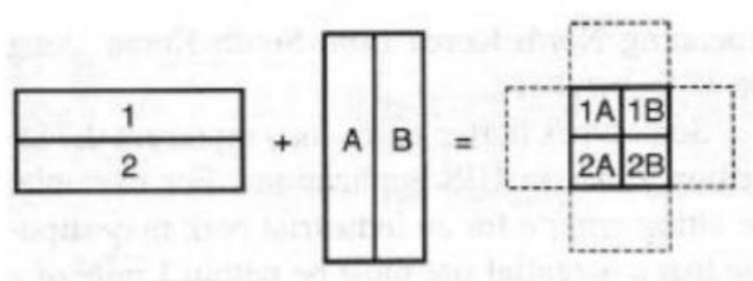


Figure 3

Overlay combines the geometries and attributes from two layers into a single layer. The dashed lines are for illustration only and are not included in the output.

- Each feature on the output contains a combination of attributes from the input layers, and this combination differs from its neighbors.

Feature Type and Overlay

- There are two groups of overlay operations.
- The first group uses two polygon layers as inputs.
- The second group uses one polygon layer and another layer, which may contain points or lines.
- Overlay operations can therefore be classified by feature type into point-in-polygon, line-in-polygon, and polygon-on-polygon.
- To distinguish the layersin the following discussion, the layer that may be a point, line, or polygon layer is called the input layer and the layer that is a polygon layer is called the overlay layer.
- In a **point-in-polygon** overlay operation the same point features in the input layer are included in the output but each point is assigned with attributes of the polygon within which it falls (Figure 4).

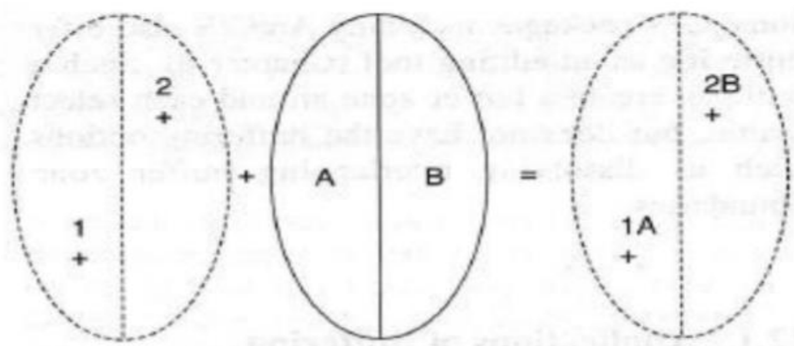


Figure 4
Point-in-polygon overlay. The input is a point layer. The output is also a point layer but has attribute data from the polygon layer.

- In a **line-in-polygon overlay** operation, the output contains the same line features as in the input layer but each line feature is dissected by the polygon boundaries on the overlay layer (Figure 6).
- Thus the output has more line segments than does the input layer. Each line segment on the output combines attributes from the input layer and the underlying polygon.
- For example, a line-in-polygon overlay can find soil data for a proposed road. The input layer includes the proposed road. The overlay layer contains soil polygons. And the output shows a dissected proposed road, each road segment having a different set of soil data from its adjacent segments.

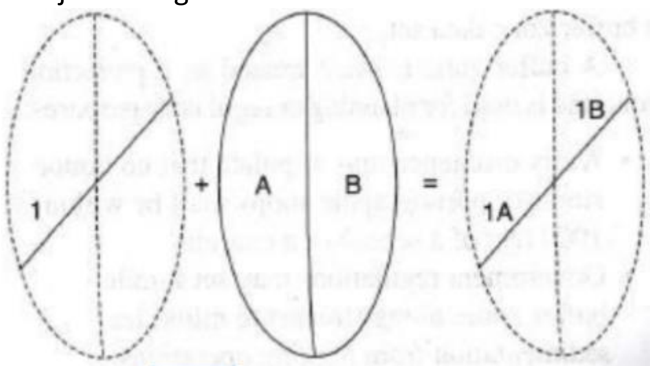


Figure 6
Line-in-polygon overlay. The input is a line layer. The output is also a line layer. But the output differs from the input in two aspects: the line is broken into two segments, and the line segments have attribute data from the polygon layer.

The most common overlay operation is polygon-on-polygon, involving two polygon layers. The output combines the polygon boundaries from the input and overlay layers to create a new set of polygons (Figure 7). Each new polygon carries attributes from both layers, and these attributes differ from those of adjacent polygons. For example, a polygon-on-polygon overlay can analyze the association between elevation zones and vegetation types.

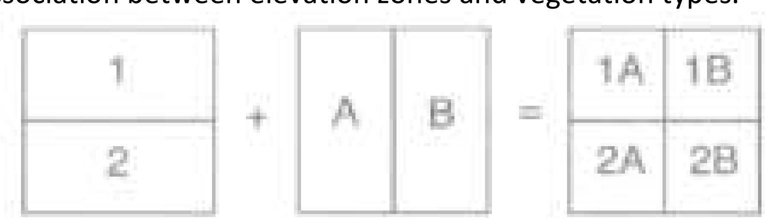


Figure 7
Polygon-on-polygon overlay. In the illustration, the two layers for overlay have the same area extent. The output combines the geometries and attributes from the two layers into a single polygon layer.

Q.5 Explain Overlay methods?

(A) all overlay methods are based on the Boolean connectors AND, OR, and XOR. An overlay operation is called Intersect if it uses the AND connector.

- An overlay operation is called Union if it uses the OR connector.

- An overlay operation that uses the XOR connector is called Symmetrical Difference or Difference.
- And an overlay operation is called Identity or Minus if it uses the following expression: [(input layer) AND (identity layer)] OR (input layer).

Union preserves all features from the inputs (Figure 8). The area extent of the output combines the area extents of both input layers. Union requires that both input layers be polygon layers.

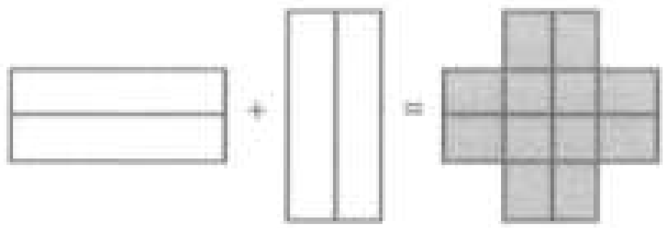


Figure 8:- The Union method keeps all areas of the two input layers in the output.

Intersect preserves only those features that fall within the area extent common to the inputs (Figure 9). The input layers may contain different feature types, although in most cases, one of them (the input layer) is a point, line, or polygon layer and the other (the overlay layer) is a polygon layer. Intersect is often a preferred method of overlay because any feature on its output has attribute data from both of its inputs.

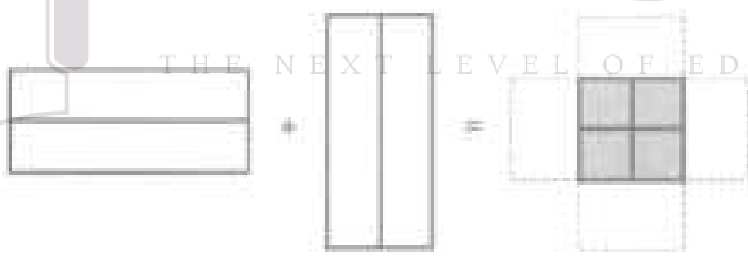


Figure 9:- The Intersect method preserves only the area common to the two input layers in the output.

Symmetrical Difference preserves features that fall within the area extent that is common to only one of the inputs (Figure 12.11). In other words, Symmetrical Difference is opposite to Intersect in terms of the output's area extent. Symmetrical Difference requires that both input layers be polygon layers.

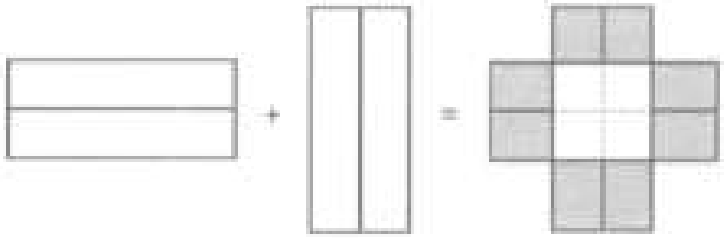


Figure 10:- The Symmetrical Difference method preserves areas common to only one of the input layers in the output.

Identity preserves only features that fall within the area extent of the layer defined as the input layer (Figure 12.12). The other layer is called the identity layer. The input layer may contain points, lines, or polygons, and the identity layer is a polygon layer.

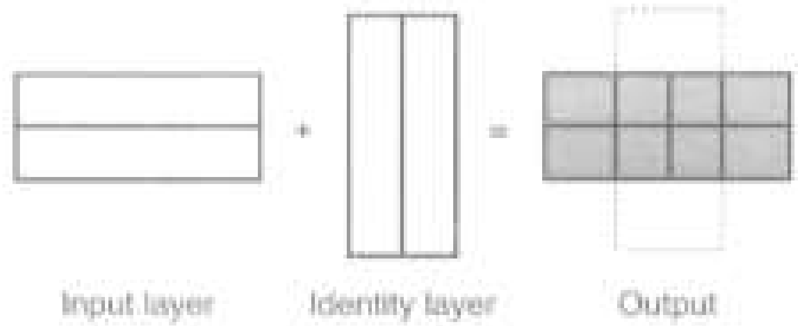


Figure 11:- The Identity method produces an output that has the same extent as the input layer. But the output includes the geometry and attribute data from the identity layer.

Q.6. What is Overlay of Shapefiles and what is its common error(silvers)

- Unlike the coverage model, both shapefile and geodatabase allow polygons to have multiple components, which may also overlap with one another.
- This means that overlay operations can actually be applied to a single feature layer: Union creates a new feature by combining different polygons, and Intersect creates a new feature from the area where polygons overlap.
- But when used with a single layer, Union and Intersect are basically editing tools for creating new features.
- Many shapefile users are aware of a problem with the overlay output: the area and perimeter values are not automatically updated.

Silvers

- A common error from overlaying polygon layers is slivers, very small polygons along correlated or shared boundary lines (e.g., the study area boundary) of the input layers (Figure 12).
- The existence of slivers often results from digitizing errors. Because of the high precision of manual digitizing or scanning, the shared boundaries on the input layers are rarely on top of one another.
- When the layers are overlaid, the digitized boundaries intersect to form slivers.
- Other causes of slivers include errors in the source map or errors in interpretation. Polygon boundaries on soil and vegetation maps are usually interpreted from field survey data, aerial photographs, and satellite images.

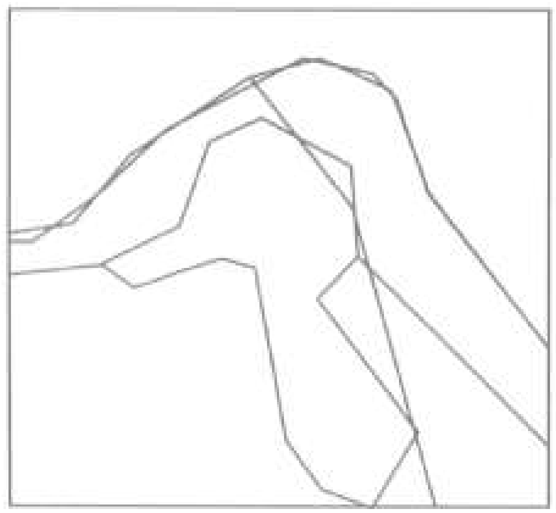


Figure 12 The top boundary has a series of slivers. These slivers are formed between the coastlines from the input layers in overlay.

- Most GIS packages incorporate some kind of tolerance in overlay operations to remove slivers.
- For example, uses the cluster tolerance, which forces points and lines to be snapped together if they fall within the specified distance (Figure 13).
- The cluster tolerance is either defined by the user or based on a default value.
- Slivers that remain on the output of an overlay operation are those beyond the cluster tolerance.
- Therefore, one option to reduce the sliver problem is to increase the cluster tolerance.

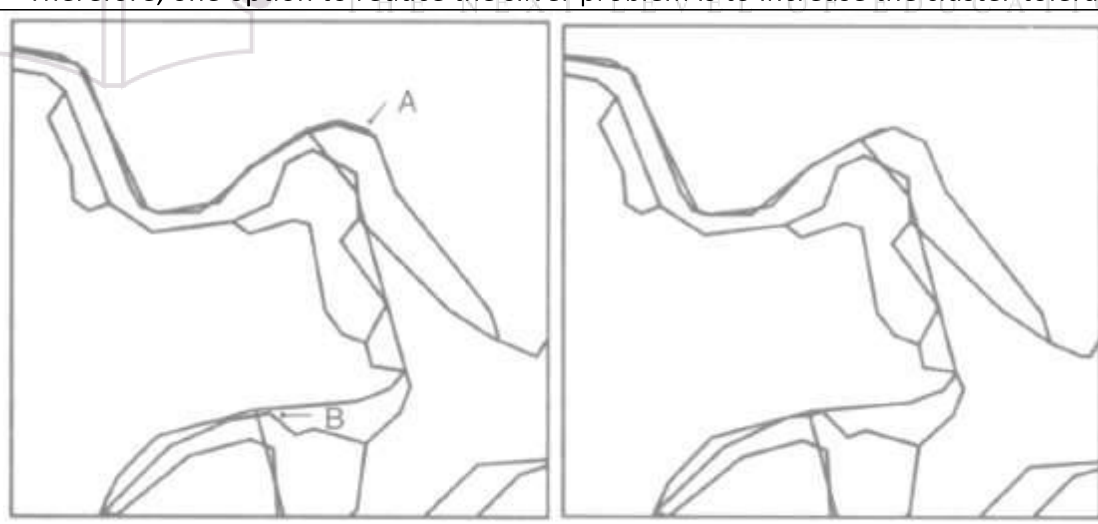


Figure 13
A cluster tolerance can remove many slivers along the top boundary (A) but can also snap lines that are not slivers (B).

Q) Explain Error Propagation in Overlay.

Error Propagation in Overlay

- Slivers are examples of errors in the inputs that can propagate to the analysis output.
- Error propagation refers to the generation of errors that are due to inaccuracies of the input layers.
- Error propagation in overlay usually involves two types of errors: positional and identification.
- Positional errors can be caused by the inaccuracies of boundaries that are due to digitizing or interpretation errors.
- Identification errors can be caused by the inaccuracies of attribute data such as the incorrect coding of polygon values.
- Every overlay product tends to have some combinations of positional and identification errors.
- How serious can error propagation be? It depends on the number of input layers and the spatial distribution of errors in the input layers.
- The accuracy of an overlay output decreases as the number of input layers increases. And the accuracy decreases if the likelihood of errors occurring at the same locations in the input layers decreases.
- An error propagation model proposed by Newcomer and Szajgin calculates the probability of the event that the inputs are correct on an overlay output. The model suggests that the highest accuracy that can be expected of the output is equal to that of the least accurate layer among the inputs, and the lowest accuracy is equal to:

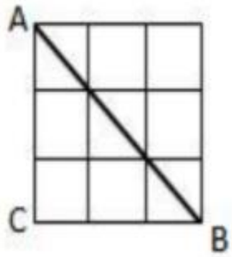
$$1 - \sum_{i=1}^{n} \Pr(Ei')$$

- where n is the number of input layers and
- Pr(Ei') is the probability that the input layer i is incorrect.

Q) DISTANCE MEASUREMENT

- Distance measurement refers to measuring straight-line (Euclidean) distances between features.
- Measurements can be made from points in a layer to points in another layer or from each point in a layer to its nearest point or line in another layer.
- Distance measures can be used directly for data analysis. For example use distance measures to test whether deer relocation points are closer to old-growth/clear-cut edges than random points located within the deer's relocation area.

a. Euclidean distance method, A straight line is drawn joining the two points and a right angled triangle is created. The distance is then derived using the Pythagorean geometry.



$$AB = \sqrt{AC^2 + CB^2}$$

If the cell size of the raster is 10m, then distance $AB = \sqrt{30^2 + 30^2}$

$$= \sqrt{1800}$$

$$= 42.42 \text{ m}$$

b. Manhattan distance method: In this method, distance along the raster cell sides from one point to the other is taken. The following example illustrates the method.

If the cell size of the raster is 10 m, then distance AB = (Aa + ab + bc + cd + de + eB)

$$= (10 + 10 + 10 + 10 + 10 + 10)$$

$$= 60 \text{ m}$$

Q) Map Manipulation

Tools are available in a GIS package for manipulating and managing maps in a database. Like buffering and overlay, these tools are considered basic GIS tools often needed for data preprocessing and data analysis. Map manipulation is easy to follow graphically, even though terms describing the various tools may differ between GIS packages.

1. Dissolve

- Aggregates features that have the same attribute value or values (Figure 14). For example, we can aggregate roads by highway number or counties by state.
- An important application of Dissolve is to simplify a classified polygon layer.
- Classification groups values of a selected attribute into classes and makes obsolete boundaries of adjacent polygons, which have different values initially but are now grouped into the same class.
- Dissolve can remove these unnecessary boundaries and creates a new, simpler layer with the classification results as its attribute values.
- Another application is to aggregate both spatial and attribute data of the input layer.
- For instance, to dissolve a county layer, we can choose state name as the attribute to dissolve and county population to aggregate.
- The output is a state layer with an attribute showing the state population (i.e., the sum of county populations).

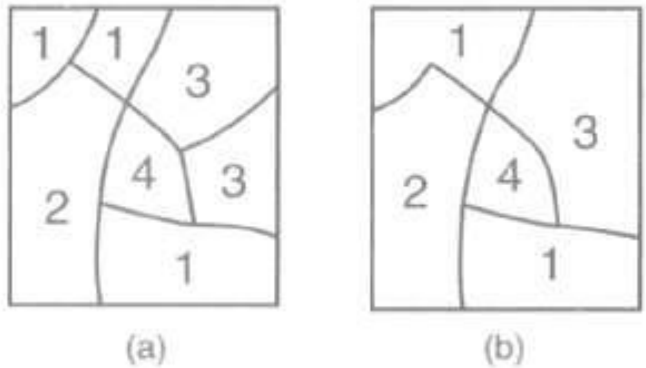


Figure 14 Dissolve removes boundaries of polygons that have the same attribute value in (a) and creates a simplified layer *(b).*

2. Clip

- creates a new layer that includes only those features of the input layer that fall within the area extent of the clip layer (Figure 15).
- Clip is a useful tool, for example, for cutting a map acquired elsewhere to fit a study area.
- The input may be a point, line, or polygon layer, but the clip layer must be a polygon layer.

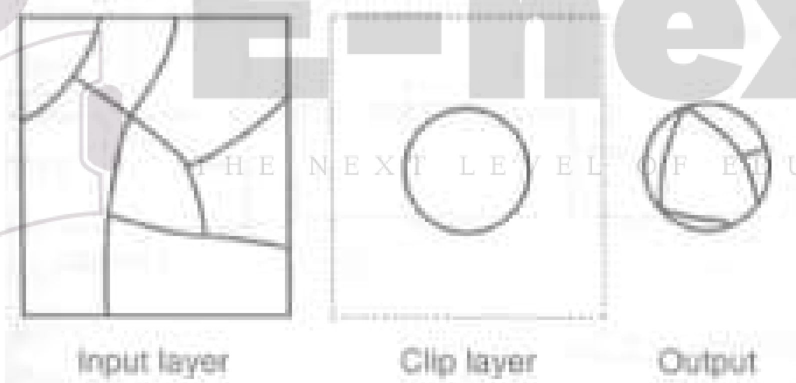


Figure 15 Clip creates an output that contains only those features of the input layer that fall within the area extent of the clip layer.

3. Append

- creates a new layer by piecing together two or more layers (Figure 12.23).
- For example, Append can put together a layer from four input layers, each corresponding to the area extent of a USGS 7.5- minute quadrangle.
- The output can then be used as a single layer for data query or display.
- But the boundaries separating the inputs still remain on the output and divide a feature into separate features if the feature crosses the boundary.

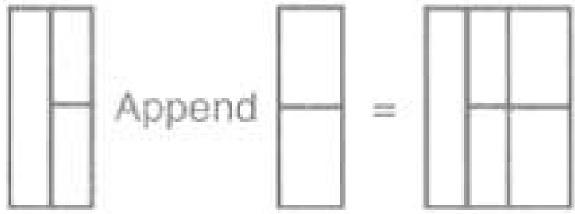


Figure 16 Append pieces together two adjacent layers into a single layer but does not remove the shared boundary between the layers.

4. Select

- creates a new layer that contains features selected from a user-defined query expression (Figure 17).
- For example, we can create a layer showing high-canopy closure by selecting stands that have 60 to 80 percent closure from a stand layer.

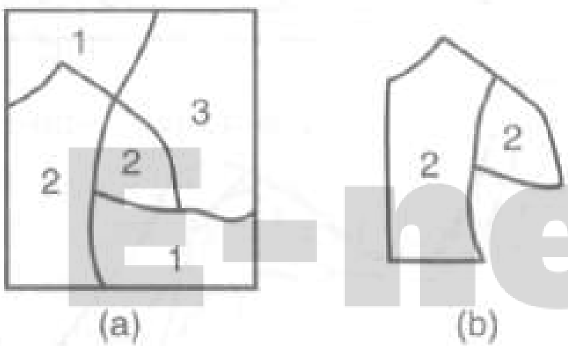


Figure 17 Select creates a new layer *(b)* with selected features from the input layer *(a)*.

5. Eliminate

- creates a new layer by removing features that meet a user-defined query expression (Figure 18).
- For example, Eliminate can implement the minimum mapping unit concept by removing polygons that are smaller than the defined unit in a layer.

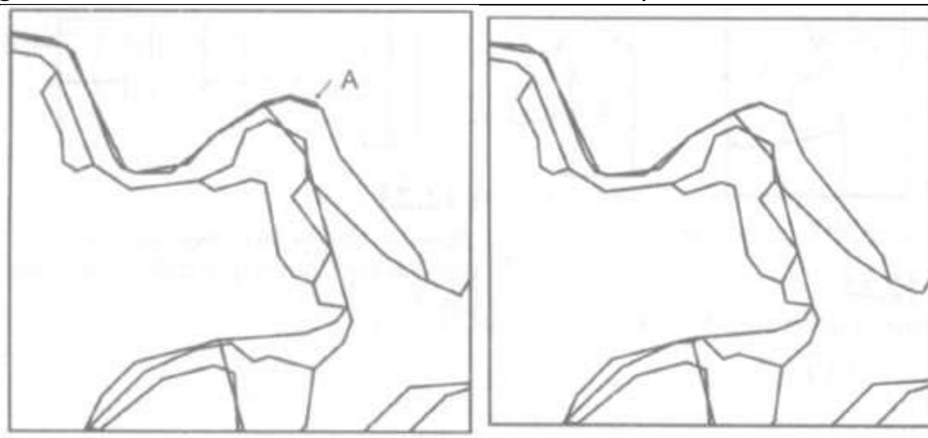


Figure 18 Eliminate removes some small slivers along the top boundary (A).

6. Update

- uses a “cut and paste” operation to replace the input layer with the update layer and its features (Figure 19).
- As the name suggests, Update is useful for updating an existing layer with new features in limited areas. It is a better option than redigitizing the entire map.

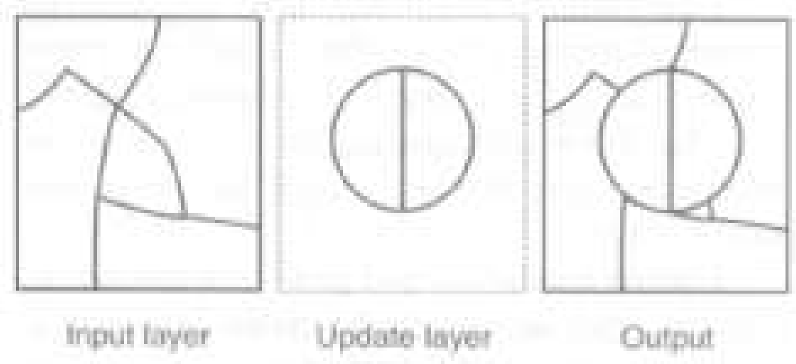


Figure 19 Update replaces the input layer with the update layer and its features.

7. Erase

- Removes from the input layer those features that fall within the area extent of the erase layer (Figure 20).
- Suppose a suitability analysis stipulates that potential sites cannot be within 300 meters of any stream.
- A stream buffer layer can be used in this case as the erase layer to remove itself from further consideration.

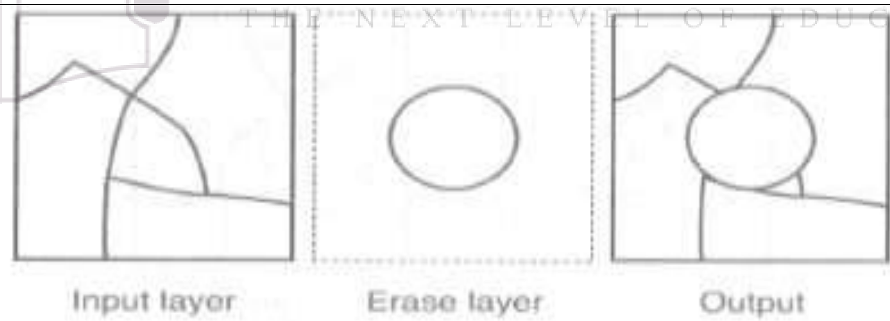


Figure 20 Erase removes features from the input layer that fall within the area extent of the erase layer.

8. Split

- Divides the input layer into two or more layers (Figure 21).
- A split layer, which shows area subunits, is used as the template for dividing the input layer.
- For example, a national forest can' split a stand layer by district so that each district office can have its own layer. In ArcGIS, Clip and Split are also editing tools.
- These editing tools work with features rather than layers.

- For example, the editing tool of Split splits a line at a specified location or a polygon along a line sketch. The tool does not work with layers.
- It is therefore important that we understand the function of a tool before using it.

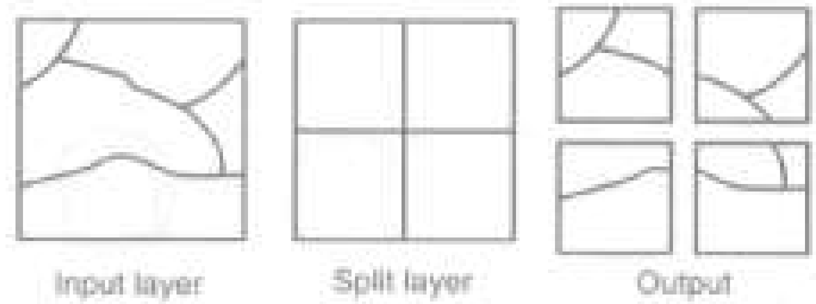


Figure 21 Split uses the geometry of the split layer to divide the input layer into four separate layers.

Q) Explain Raster data analysis and data analysis environment

RASTER DATA ANALYSIS

- The raster data model uses a regular grid to cover the space and the value in each grid cell to represent the characteristic of a spatial phenomenon at the cell location.
- This simple data structure of a raster with fixed cell locations not only is computationally efficient, but also facilitates a large variety of data analysis operations.
- In contrast to vector data analysis, which is based on the geometric objects of point. line, and polygon, raster data analysis is based on cells and rasters.
- Raster data analysis can be performed at the level of individual cells, or groups of cells. Or cells within an entire raster.
- Some raster data operations use a single raster others use two or more rasters.
- An important consideration in raster data analysis is the type of cell value. Statistics such as mean and standard deviation are designed for numeric values, whereas others such as majority (the most frequent cell value) are designed for both numeric and categorical values.

DATA ANALYSIS ENVIRONMENT

- The analysis environment refers to the area for analysis and the output cell size.
- The area extent for analysis may correspond to a specific raster, or an area defined by its minimum and maximum x-, y-coordinates, or a combination of rasters.
- Given a combination of rasters with different area extents, the area extent for analysis can be based on the union or intersect of the rasters.
- The union option uses an area extent that encompasses all input rasters, whereas the intersect option aces an area extent that is common to all input rasters.
- An analysis mask can also determine the area extent for analysis.
- An analysis mask limits analysis to cells that do not carry the cell value of "no data."
- No data differs from zero. Zero is a valid cell value, whereas no data means the absence of data.

- We can define the output cell size at any scale deemed suitable. Typically, the output cell size is set to be equal to, or larger than, the largest cell size among the input rasters.
- This follows the rationale that the accuracy of the output should correspond to that of the least accurate input raster.

Q) explain Local Operations in raster data analysis.

LOCAL OPERATIONS

- Constituting the core of raster data analysis, local operations are cell-by-cell operations.
- A local operation can create a new raster from either a single input raster or multiple input raster's.
- The cell values of the new raster are computed by a function relating the input to the output or are assigned by a classification table.

Local Operations with a Single Raster

- Given a single raster as the input, a local operation computes each cell value in the output raster as a mathematical function of the cell value in the input raster.
- A large number of mathematical functions are available in a GIS package.(Figure 22)
- Converting a floating-point raster to an integer raster, for example, is a simple local operation that uses the integer function to truncate the cell value at the decimal point on a cell-by-cell basis.
- Converting a slope raster measured in percent to one measured in degrees is also a local operation but requires a more complex mathematical expression.
- In Figure 23, the expression [slope_d] = 57.296 X arctan ([slope_p]/100) can convert *slope_p* measured in percent to *slope_d* measured in degrees.
- Because computer packages typically use radian instead of degree in trigonometric functions, the constant 57.296 (360/2π, π = 3.1416) changes the angular measure to degrees.

Arithmetic	+, -, /, *, absolute,integer,floating-point
Logarithmic	exponentials, logarithms
Trigonometric	sin, cos, tan, arcsin, arccos, arctan
Power	square, square root, power

Figure 22 Arithmetic, logarithmic, trigonometric, and power functions for local operations.

15.2	16.0	18.5
17.8	18.3	19.6
18.0	19.1	20.2

(a)

8.64	9.09	10.48
10.09	10.37	11.09
10.20	10.81	11.42

(b)

Figure 23 A local operation can convert a slope raster from percent *(a)* to degrees (*b*).

Reclassification

A local operation, reclassification creates a new raster by classification. Reclassification is also referred to as recoding, or transforming, through lookup tables (Tomlin 1990). Two reclassification methods may be used.

- The first method is a one-to-one change, meaning that a cell value in the input raster is assigned a new value in the output raster.
- The second method assigns a new value to a range of cell values in the input raster.

Local Operations with Multiple Rasters

- Local operations with multiple rasters are also referred to as compositing, overlaying, or superimposing maps (Tomlin 1990).
- Another common term for local operations with multiple input rasters is map algebra, a term that refers to algebraic operations with raster map layers (Tomlin 1990; Pullar 2001).
- Because local operations can work with multiple rasters, they are the equivalent of vector-based overlay operations.
- A greater variety of local operations have multiple input rasters than have a single input raster.
- Besides mathematical functions that can be used on individual rasters, other measures that are based on the cell values or their frequencies in the input rasters can also be derived and stored on the output raster.
- Some of these measures are, however, limited to rasters with numeric data.
- Summary statistics, including maximum, minimum, range, sum, mean, median, and standard deviation, are measures that apply to rasters with numeric data.
- Figure 24, for example, shows a local operation that calculates the mean from three input rasters.
- If a cell contains no data in one of the input rasters, the cell also carries no data in the output raster by default.

- Other measures that are suitable for rasters with numeric or categorical data are statistics such as majority, minority, and number of unique values.
- For each cell, a majority output raster tabulates the most frequent cell value among the input rasters, a minority raster tabulates the least frequent cell value, and a variety raster tabulates the number of different cell values.
- Figure 25, for example, shows the output with the majority statistics from three input rasters.

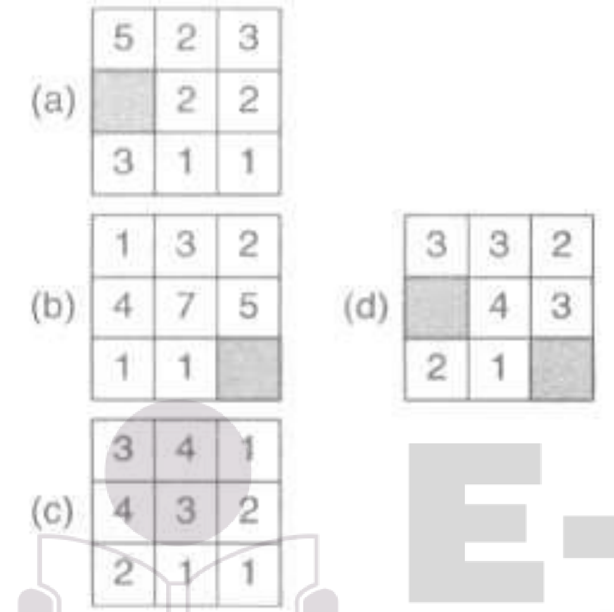


Figure 24 The cell value in *(d)* is the mean calculated from three input rasters *(a. b,* and c) in a local operation. The shaded cells have no data.

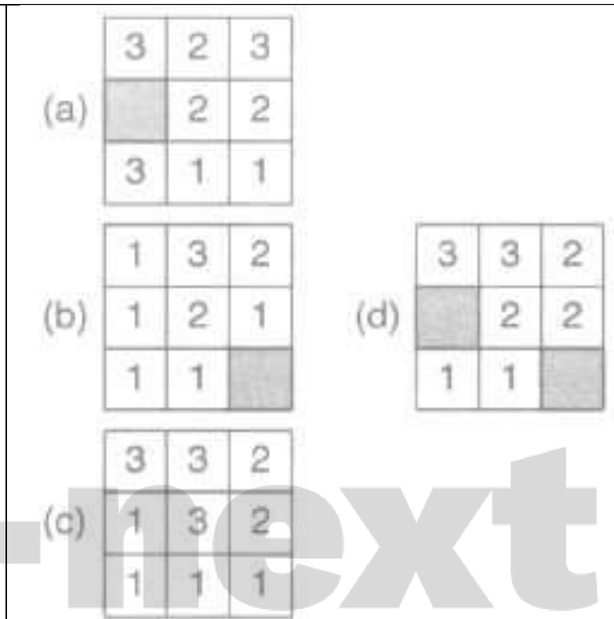


Figure 25 The cell value in *(d)* is the majority statistic derived from three input rasters *(a, b,* and c) in a local operation. The shaded cells have no data.

- Some local operations do not involve statistics or computation.
- A local operation called Com-bine assigns a unique output value to each unique combination of input values.
- Suppose a slope raster has three cell values (0 to 20%, 20 to 40%, and greater than 40% slope), and an aspect raster has four cell values (north, east, south, and west aspects).
- The Combine operation creates an output raster with a value for each unique combination of slope and aspect, such as 1 for greater than 40% slope and the south aspect, 2 for 20 to 40% slope and the south aspect and so on (Figure 24).

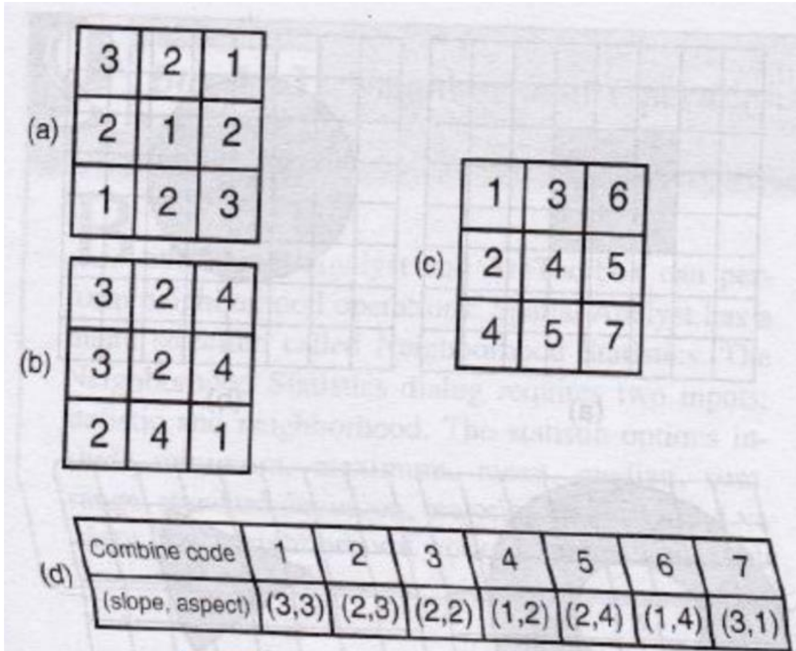


Figure 24
Each cell value in (*c*) represents a unique combination of cell values in (*a*) and (*b*). The combination codes and their representations are shown in (*d*).

Q) explain neighborhood operation and its applications?

NEIGHBORHOOD OPERATIONS

- A neighborhood operation involves a focal cell and a set of its surrounding cells.
- The surrounding cells are chosen for their distance and/or directional relationship to the focal cell.
- Common neighborhoods include rectangles, circles, annuluses, and wedges (Figure 25).
- A rectangle is defined by its width and height in cells, such as a 3-by-3 area centered at the focal cell.
- A circle extends from the focal cell with a specified radius.
- An annulus or doughnut-shaped neighborhood consists of the ring area between a smaller circle and a larger circle centered at the focal cell.
- And a wedge consists of a piece of a circle centered at the focal cell.
- As shown in Figure 5.6 some cells are only partially coveted in the defined neighborhood. The general rule is to include a cell if the center of the cell falls within the neighborhood.

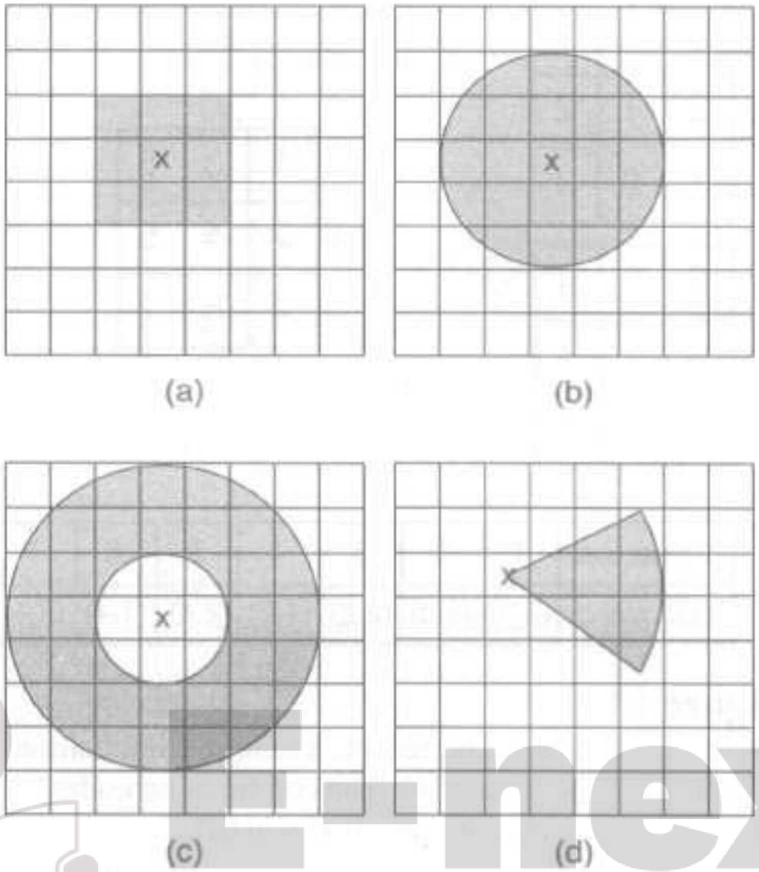


Figure 26 Four common neighborhood types: rectangle (*a*), circle(,*b*), annulus (c), and wedge *(d)*. The cell marked with an x is the focal cell.

Neighborhood Statistics

- A neighborhood operation typically uses the cell values within the neighborhood, with or without the focal cell value, in computation, and then as-signs the computed value to the focal cell.
- To complete a neighborhood operation on a raster, the focal cell is moved from one cell to another until all cells are visited.
- Although a neighborhood operation works on a single raster, its process is similar to that of a local operation with multiple rasters.
- Instead of using cell values from different input rasters a neighborhood operation uses the cell values from a defined neighborhood.
- The output from a neighborhood operation can show summary statistics including maximum, minimum, range, sum, mean, median, and standard deviation, as well as tabulation of measures such as majority, minority, and variety.
- These statistics and measures are the same as those from local operations with multiple rasters.
- A block operation is a neighborhood operation that uses a rectangle (block) and assigns the calculated value to all block cells in the output raster.

- Therefore, a block operation differs from a regular neighborhood operation because it does not move from cell to cell but from block to block.

Applications of Neighborhood Operations

- The moving **average method**, for instance, reduces the level of cell value fluctuation in the input raster (Figure 27).
- Edge enhancement, for example, can use a range filter, essentially a neighborhood operation using the **range statistic** (Figure 28).
- The range measures the difference between the maximum and minimum cell values within the defined neighborhood.
- The opposite of edge enhancement is a smoothing operation that is based on the (Figure 29).
- The majority operation assigns the most frequent cell value within the neighborhood to the focal cell, thus creating a smoother raster than the original raster.

(a)

1	2	2	2	2
1	2	2	2	3
1	2	1	3	3
2	2	2	3	3
2	2	2	2	3

(b)

1.56	2.00	2.22
1.67	2.11	2.44
1.67	2.11	2.44

Figure 27 The cell values in *(b)* are the neighborhood means of the shaded cells in *(a)* using a 3-by-3 neighborhood. For example, 1.56 in the output raster is calculated from **(1 + 2 + 2 + 1 + 2 + 2 + 1 + 2 + 1)/9**.

(a)

200	200	110	210	210
200	200	110	210	210
150	150	100	170	170
140	140	130	160	160
140	140	130	160	160

(b)

100	110	110
100	110	110
50	70	70

Figure 28 The cell values in (*b*) are the neighborhood range statistics of the shaded cells in *(a)* using a 3-by-3 neighborhood. For example, the upper-left cell in the output raster has a cell value of 100, which is calculated from (200 — 100).

(a)

1	2	2	2	2
1	2	2	2	3
1	2	1	3	3
2	2	2	3	3
2	2	2	2	3

(b)

2	2	2
2	2	3
2	2	3

Figure 29
The cell values in (*b*) are the neighborhood majority statistics of the shaded cells in *(a)* using a 3-by-3 neighborhood. For example, the upper-left cell in the output raster has a cell value of 2 because there are five 2s and four 1s in its neighborhood.

Q) explain zonal operations and zonal statistics?

Zonal Operations

- A zonal operation works with groups of cells of same values or like features.
- These groups are called zones.
- Zones may be contiguous or noncontiguous.
- A contiguous zone includes cells that are spatially connected, whereas a noncontiguous zone includes separate regions of cells.

Zonal Statistics

- A zonal operation may work with a single raster or two rasters.
- Given a single input raster, zonal operations measure the geometry of each zone in the raster, such as area, perimeters, thickness, and centroid.(Figure 29)
- The area is the sum of the cells that fall within the zone times the cell size.
- The perimeter of a contiguous zone is the length of its boundary and the perimeter of a noncontiguous zone is the sum of the length of each region.
- The thickness calculates the radius (in cells) of the largest circle that can be drawn within each zone. And the centroid is the geometric center of a zone located at the intersection of the major axis and the minor axis of an ellipse that best approximates the zone.
- Given two rasters in a zonal operation, one in-put raster and one zonal raster, a zonal operation produces an output raster, which summarizes the cell values in the input raster for each zone in the zonal raster.
- The summary statistics and measures include area, minimum, maximum, sum, range. mean, standard deviation, median, majority, minority and variety. (The last four measures are not available if the input raster is a floating-point raster.)
- Figure 30 shows a zonal operation of computing the mean by zone.

Figure 29
Thickness and centroid for two large watersheds (zones). Area is measured in square kilometers, and perimeter and thickness are measured in kilometers. The centroid of each zone is marked with an x.

Zone	Area	Perimeter	Thickness
1	36,224	1708	77.6
2	48,268	1464	77.4

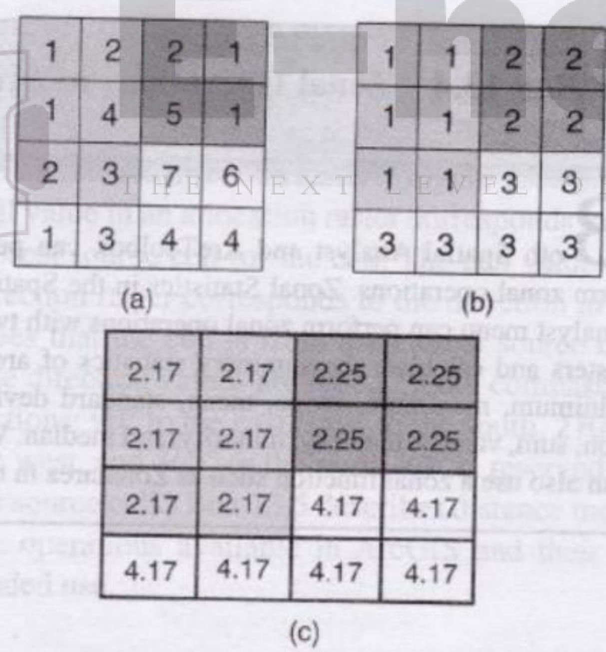


Figure 30
The cell values in (*c*) are the zonal means derived from an input raster (*a*) and a zonal raster (*b*). For example, 2.17 is the mean of {1, 1, 2, 2, 4, 3} for zone 1.

 Figure 30b is the zonal raster with three zones. Figure 30a is the input raster and Figure 30c is the output raster.

Q) Explain physical distance measures operations

PHYSICAL DISTANCE MEASURE OPERATIONS

- In a GIS project, distances may be expressed as physical distances or cost distances.
- The physical distance measures the straight-line or Euclidean distance, whereas the cost distance measures the cost for traversing the physical distance.
- The distinction between the two types of distance measures is important in real-world applications.
- A truck driver, for example, is more interested in the time or the fuel cost for covering a route than in its physical distance.
- The cost distance in this case is based on not only the physical distance but also the speed limit and road condition.
- This section focuses on the physical distance.

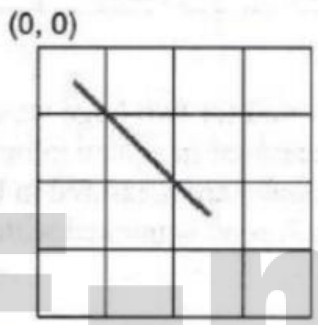


Figure 31
A straight-line distance is measured from a cell center to another cell center. This illustration shows the straight-line distance between cell (1,1) and cell (3,3).

- Physical distance measure operations calculate straight-line distances away from cells designated as the source cells. For example, to get the distance between cells (1,1) and (3,3) in Figure 31, we can use the following formula:

$$\textbf{cell size} * \sqrt{(3-1)2 + (3-1)^2}$$

- A physical distance measure operation essentially buffers the source cells with wavelike continuous distances over the entire raster (Figure 32) or to a specified maximum distance.

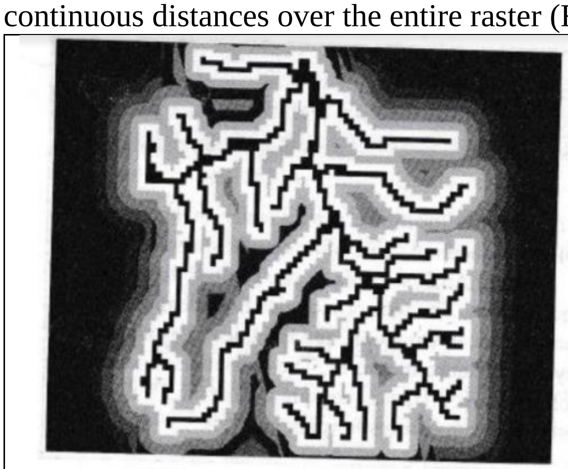


Figure 32
Continuous distance measures from a stream network.

- This is why physical distance measure operations are also extended neighborhood operations or global operations.
- This option is based on the consideration of convenience because the data set is converted from vector to raster data before the operation starts.
- The continuous distance raster from a physical distance measure operation can be used directly in subsequent operations.
- But more often it is further processed to create a specific distance zone, or a series of distance zones from the source cells.
- Reclassify can convert a continuous distance raster into a raster with one or more discrete distance zones.
- A variation of Reclassify is an operation called Slice, which can divide a continuous distance raster into equal-interval or equal-area distance zones.

Q) Explain Allocation and Direction. Write down Applications of Physical Distance Measure Operations

- Besides calculating straight-line distances, physical distance measure operations can also produce allocation and direction rasters (Figure 33).

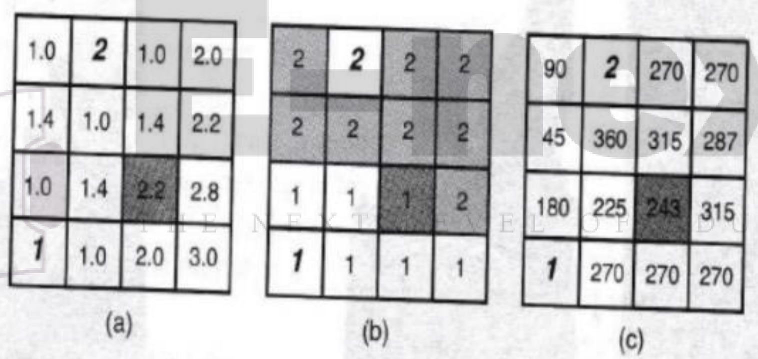


Figure 33
Based on the source cells denoted as *1* and *2*, (*a*) shows the physical distance measures in cell units from each cell to the closest source cell; (*b*) shows the allocation of each cell to the closest source cell; and (*c*) shows the direction in degrees from each cell to the closest source cell. The cell in a dark shade (row 3, column 3) has the same distance to both source cells. Therefore, the cell can be allocated to either source cell. The direction of 243° is to the source cell 1.

- The cell value in an allocation raster corresponds to the closest source cell for the cell.
- The cell value in a direction raster corresponds to the direction in degrees that the cell is from the closest source cell.
- The direction values are based on the compass directions: 90° to the east, 180° to the south, 270° to the west and 360° to the north.

Applications of Physical Distance Measure Operations

- Like buffering around vector-based features, physical distance measure operations have many applications.
- For example, we can create equal-interval distance zones from a stream network or regional fault lines.